The Scientific Method

A Guide to Finding Useful Knowledge

J. Scott Armstrong
Wharton School, University of Pennsylvania

Kesten C. Green
University of South Australia

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

103 Penang Road, #05-06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781316515167
DOI: 10.1017/9781009092265

First published 2022

A catalogue record for this publication is available from the British Library.

ISBN 978-1-316-51516-7 Hardback
ISBN 978-1-009-09642-3 Paperback

3 CHECKLIST FOR THE SCIENTIFIC METHOD

We intend that our checklist provides a common understanding among all stakeholders in science of what the scientific method entails. To that end, we describe it in terms that are simple and commonly understood.

In this chapter, we outline how we developed the *Compliance with Science Checklist*. We then present the checklist of eight criteria for complying with the scientific method and 26 items to help check whether the criteria are met. The checklist is intended for all stakeholders of science. We describe how the checklist can be used, and list stakeholders and what they can use the checklist for in Table 3.1. We caution that checklists are only useful if they are logical and based on evidence, and if they are used.

3.1 Development of the *Compliance With Science Checklist*

Checklists draw upon the decomposition principle, which reduces a complex problem into simpler parts. One solves or makes estimates for, or rates, each part, and then calculates an aggregate solution, or overall rating.

Our review of experimental evidence showed that decomposition typically provides substantial improvements in predictive validity. For example, in three experiments on subjects' decisions for job and college selection, judgmental decomposition resulted in more accurate judgments than holistic ratings (Arkes et al., 2010). Similarly, an experiment in which members of the Society for Medical Decision Making

https://doi.org/10.1017/9781009092265.004 Published online by Cambridge University Press

Table 3.1. Potential users and uses of the *Compliance With Science Checklist*

Researchers	▪ determining which findings to cite ▪ ensuring that their own papers comply ▪ informing clients, editors, users, and readers on the extent to which their paper complies
Journals	▪ setting expectations of authors ▪ identifying which criteria were met ▪ selecting which papers to publish
Universities	▪ training, hiring, promoting, and dismissing scientists ▪ setting expectations of researchers ▪ disseminating useful scientific findings
Think Tanks	▪ assessing papers to identify the scientific criteria that were met
Funders	▪ requiring research to meet scientific criteria
Awards Committees	▪ choosing recipients who made useful scientific discoveries
Certifiers	▪ independently assessing the extent to which papers provide useful scientific findings
Managers	▪ assessing the value of published findings
Journalists	▪ reporting the extent to which studies address important problems and comply with science
Regulators	▪ developing, revising, and rescinding regulations based on compliance with science
Law Courts	▪ assessing the value of evidence

evaluated presentations at their annual convention found that decomposed ratings were more reliable than holistic ratings (Arkes et al., 2006). For additional experimental evidence on the value of decomposition, see MacGregor (2001).

To develop a checklist of criteria for compliance with the scientific method, we reviewed experimental research on scientific practice (described in Chapter 4). Based on the research findings, we designed *operational guidelines* for each of the eight criteria. For example, to gauge a paper's objectivity, the checklist asks raters to determine whether a paper compares multiple reasonable hypotheses.

As we will show in this book, the *Compliance With Science Checklist* provides a valid and reliable way to rate the extent to which

https://doi.org/10.1017/9781009092265.004 Published online by Cambridge University Press

papers – or methods or policies – comply with the scientific method. This checklist, along with other checklists in this book, is also provided at GuidelinesForScience.com. To ensure that raters understood the guidelines, we pretested the checklist many times by examining the inter-rater reliability of the ratings for each of the criteria.

Checklist 3.1 is the result of our efforts: it provides 26 operational items to rate compliance with the eight criteria of the scientific method.

Checklist 3.1 *Compliance With Science Checklist*

Paper title:		
Reviewer:	**Date:**	**Time spent (minutes):**

Instructions for Raters

1. Skim the paper while you complete the checklist ***as a skeptical reviewer***.
2. **Rate each lettered item**, below, marking the relevant checkbox to indicate
   True if the research complies,
   F/? (False/Unclear) if the research does *not* comply, or if you are unsure.
   IMPORTANT: If you are ***not convinced*** that the paper complied, rate the item **F/?**
3. If you rate an item **True**, ***give reasons for your rating in your own words***.
4. **Rate criteria 1–8** as **True** by marking the checkbox only if all lettered items for the criterion are rated **T**.

First assess whether the paper complies with the lettered items under each criterion below. Then assess whether it complies with each of the eight criteria based on compliance with the lettered items. ***Avoid speculation.***	
1. Problem is important for decision-making, policy, or method development	☐ **True** **T** **F/?**
a. Importance of the problem clear from the title, abstract, result tables, or conclusions	☐ ☐
b. Findings add to cumulative scientific knowledge	☐ ☐
c. Uses of the findings are clear to you	☐ ☐
d. The findings can be used to improve people's lives without resorting to duress or deceit	☐ ☐
2. Prior knowledge was comprehensively reviewed and summarized	☐ **True** **T** **F/?**
a. The paper describes objective and comprehensive procedures used to search for prior useful scientific knowledge	☐ ☐

https://doi.org/10.1017/9781009092265.004 Published online by Cambridge University Press

Checklist 3.1 cont'd

b. The paper describes how prior substantive findings were used to develop hypotheses (e.g. direction and magnitude of effects of causal variables) and research procedures	☐ ☐
3. Disclosure is sufficiently comprehensive for understanding and replication	☐ **True** **T F/?**
a. Methods are fully and clearly described so as to be understood by all relevant stakeholders, including potential users	☐ ☐
b. Data are easily accessible using information provided in the paper	☐ ☐
c. Sources of funding are described, or absence of external funding noted	☐ ☐
4. Design is objective (*unbiased by advocacy*)	☐ **True** **T F/?**
a. Prior hypotheses are clearly described (e.g., regarding directions and magnitudes of relationships, and effects of conditions)	☐ ☐
b. All reasonable hypotheses are included in the design, including plausible naive, no-meaningful-difference, and current-practice hypotheses	☐ ☐
c. Revisions to hypotheses are described, or absence of revisions noted	☐ ☐
5. Data are valid (true measures) and reliable (repeatable measures)	☐ **True** **T F/?**
a. Data were shown to be relevant to the problem	☐ ☐
b. All relevant data were used, including the longest relevant time-series	☐ ☐
c. Reliability of data was assessed	☐ ☐
d. Other information needed for assessing the validity of the data is provided, such as adjustments, known shortcomings and potential biases	☐ ☐
6. Methods were validated (proven fit for purpose) and simple	☐ **True** **T F/?**
a. Methods were explained clearly and shown valid – unless well known to intended readers, users, and reviewers, and validity is obvious	☐ ☐

https://doi.org/10.1017/9781009092265.004 Published online by Cambridge University Press

Checklist 3.1 cont'd

b. Methods were sufficiently simple for potential users to understand	☐ ☐
c. Multiple validated methods were used	☐ ☐
d. Methods used cumulative scientific knowledge explicitly	☐ ☐
7. Experimental evidence was used to compare alternative hypotheses	☐ **True** **T F/?**
a. Experimental evidence was used to compare hypotheses under explicit conditions	☐ ☐
b. Predictive validity of hypotheses was tested using out-of-sample data	☐ ☐
8. Conclusions follow logically from the evidence presented	☐ **True** **T F/?**
a. Conclusions do not go beyond the evidence in the paper	☐ ☐
b. Conclusions are not the product of confirmation bias	☐ ☐
c. Conclusions do not reject a hypothesis by denying the antecedent	☐ ☐
d. Conclusions do not support a hypothesis by affirming the consequent	☐ ☐

Describe the most important scientific finding in your own words.

Sum the criteria (1–8) rated True for compliance: [ ] of 8

An electronic version of this checklist is available at guidelinesforscience.com.

3.2 Using the Checklist: For What, How, and by Whom

The *Compliance With Science Checklist* is intended to help researchers discover useful scientific knowledge and stakeholders to evaluate research.

As far as we are aware, the *Compliance With Science Checklist* is the only checklist designed for assessing the extent to which a paper complies with the scientific method. For example, a major US research funding body, the National Science Foundation, states that the agency was created by Congress in 1950 with a mission to "promote the

https://doi.org/10.1017/9781009092265.004 Published online by Cambridge University Press

progress of science" in its *Proposal and Award Policies and Procedures Guide* (National Science Foundation, 2019, p. viii), yet the agency does not define what it means by "science."

The *Compliance With Science Checklist* can be used for different purposes. For example, when it is used to assess whether to cite a paper for its scientific findings, researchers will find that they can typically complete the checklist in fewer than five minutes.

Researchers could also use the checklist to assess the compliance of their own papers before submission to a journal. Researchers should keep in mind that *they* are responsible for writing a paper that convinces raters that their research complied with the scientific method.

Before rating compliance with science for papers by others, raters should report potential biases, and sign an oath that: "I will rate this paper to the best of my ability and without bias." Raters who are uncomfortable signing such an oath, should not rate the paper.

An assessment of a research paper's compliance with all eight criteria takes less than half an hour on average. That estimate is based on the experiences of our research assistants, who rated more than 500 papers for us.

Anyone with a stake in useful scientific research can use the ratings from the completed *Compliance With Science Checklist*. Table 3.1 provides a list of potential users along with suggestions on how they could use the checklist.

For example, independent rating organizations could provide *Compliance with Science* ratings of papers as part of formal certification procedures, and for any of the other purposes listed in the table. Rating firms could meet the likely demand for ratings of university departments on the extent to which the research output of their researchers complies with the scientific method.

3.3 Not All Checklists Are Useful

If checklist items are irrelevant, misleading, or not based on scientific evidence or logic, the use of the checklist would be expected to harm decision-making.

Harmful checklists are often used in management. In one example, Porter (1980) proposed his "five forces" framework for competitive strategy planning based on opinions. To the best of our knowledge, the "forces" were not supported by experimental evidence, economic theory, or logic, as Rasmussen (2017) explained.

https://doi.org/10.1017/9781009092265.004 Published online by Cambridge University Press

In another example, a series of laboratory experiments tested the value of the Boston Consulting Group's "BCG matrix," a four-item checklist for selecting investment opportunities. The subjects – 1,015 management students – worked independently. They were asked to choose between an investment opportunity that would double their investment and another that would lose half of their investment. Six researchers, each from a different country, ran experiments on 27 occasions during a five-year period. Of subjects exposed to the BCG checklist, 64 percent selected the unprofitable investment. Of those who were not exposed to the BCG matrix, 45 percent selected the unprofitable investment (Armstrong and Brodie, 1994). And, yes, it is a concern that only 55 percent of "unexposed" management students selected the profitable project.

3.4 Ensuring That Checklists Are Used

Consider the ARRIVE guidelines for animal studies. The 20-item checklist of guidelines was supported, but not required, by over 300 journals and major funders. A study of papers published in *PLOS* and *Nature* journals in the two years before and after the guidelines were introduced in 2010 suggested that "authors, referees, and editors generally are ignoring guidelines" (Baker et al., 2014, p. 1).

Is it sufficient to require completion of a checklist of guidelines? In a follow-up study on compliance with the ARRIVE guidelines, authors of 332 manuscripts were sent a copy of the checklist and told that they *must* complete it for their paper to be accepted, while authors of 340 manuscripts were not. There was little difference in the usage of the checklist and, despite the requirement, the papers of authors in the "must be completed" treatment were published regardless of whether they had done so. A completed checklist was requested again from the authors in the treatment group if they failed to comply the first time. Follow-up was not effective for compliance, either (Hair et al., 2018).

We find the lack of compliance with the requirement for completing a checklist strange. As experimenters, we have little trouble in getting subjects to complete checklists. We simply make it part of the contract. In our experimental studies, the subjects have always used the checklists as directed. In short, if a client states that payment for a project will only be made if a checklist is followed, we expect that nearly all who accept the contract and who are capable of completing the task will do so. We expect that if journals insisted on a completed checklist before considering a paper, they would have similar success.